DOCUMENT RESUME

ED 354 750 FL 020 948

| | |
|---|---|
| AUTHOR | Tong, Keith; And Others |
| TITLE | To Test or Not To Test, That Is the Question. |
| REPORT NO | ISSN-1015-2059 |
| PUB DATE | 91 |
| NOTE | 10p.; For the serial issue from which this paper is analyzed, see FL 020 947. |
| PUB TYPE | Reports - Evaluative/Feasibility (142) -- Journal Articles (080) |
| JOURNAL CIT | Hongkong Papers in Linguistics and Language Teaching; v14 p15-23 1991 |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | College Freshmen; Content Validity; *English (Second Language); Factor Analysis; Foreign Countries; Higher Education; High School Graduates; *Language Proficiency; Language Research; *Language Tests; Linguistic Theory; Predictive Validity; *Student Placement; *Testing; Test Reliability; *Test Use |
| IDENTIFIERS | *Hong Kong; Use of English Examination (Hong Kong) |

ABSTRACT

A study investigated the content and predictive validity of Hong Kong's new Use of English Examination (UEE), introduced by the Hong Kong Examinations Authority in 1989. The study set out to discover whether the examination was reliable for determining students' language needs on entering higher education, and whether institutions' existing internal language tests were still necessary. The UEE was compared with the University of Hong Kong's Language Analysis Sessions (LAS) test, used for arts faculty students. Subtest scores on the two tests and end-of-year subject examination scores for 346 university freshmen were analyzed statistically, using correlation and factor analysis. Results suggest that for the purpose of predicting prospective undergraduates' academic performance, the LAS tests are more effective than the UEE. However, it is noted that the former are targeting a more homogeneous and more highly-skilled population. Despite considerable overlap in the skills tested, the UEE and LAS are considered different and not parallel, suggesting that internal institutional testing may still be useful. The new UEE is also seen as more valid than the old UEE in terms of face and predictive validity. (MSE)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

HONGKONG PAPERS IN LINGUISTICS AND LANGUAGE TEACHING 14 (1991) ISSN 1015-2059

ED354750

# TO TEST OR NOT TO TEST, THAT IS THE QUESTION

Keith Tong
Hong Kong University of Science and Technology

Rose Chan & Jo Lewkowicz
University of Hong Kong

## Introduction

In 1989 the Hong Kong Examinations Authority (HKEA) introduced the new Use of English examination. The specific objectives of the examination were to:

> ... test the ability of the candidates to understand and use English as might be required in tertiary education and/or future employment. (HKEA, 1987: 2)

The examination sets out to test candidates' ability to:

> (i) understand and interpret spoken English as it might be encountered in academic and vocational situations;
>
> (ii) write clear, concise and grammatical English in an appropriate style;
>
> (iii) demonstrate both global and detailed understanding of a variety of written texts;
>
> (iv) integrate reading, writing and study skills in the pursuance of task-based/problem solving activities. (HKEA, ibid)

The examination was therefore designed to test candidates' readiness to use English in tertiary education and employment. Its development raised the question: would the examination results give sufficient information to determine whether students entering tertiary level education needed further English programmes to facilitate their academic studies? If so, there would appear to be no need to retest students at the beginning of their university career.

The purpose of this study was to investigate the content and predictive validity of the new Use of English examination. More specifically, it set out to find out whether the examination could be relied on for determining students' language needs on entering tertiary education. Since prior to the introduction of the new UE examination institutions such as the University of Hong Kong (HKU) had developed their own placement/proficiency tests for this purpose, it was important to find out whether these internal tests would continue to be necessary.

The comprehensive battery of tests developed by the Language Centre of HKU to assess Faculty of Arts students is known as the Language Analysis Sessions (LAS). These tests cover all four language skills and are therefore time-consuming to administer and mark. However, when the tests were validated in the mid-1980's (Fok, 1985), they were found to have better predictive power than the old (pre-1989) Use of English examination in terms of students' end-of-first-year results in their academic studies. They were also found to be superior in terms of face validity, since they reflected more closely the language skills required of the students in their academic work. They were not only seen by the students but also by the academic staff of the Faculty of Arts as reflecting

FL 620 948

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

the tasks demanded of the first year students. There was therefore every justification for administering the tests.

There was also a further reason for administering the LAS test to students of the Faculty of Arts. These students were originally assigned to courses addressing specific language skill areas. For example, students found to be weak in spoken English but strong on reading and writing were assigned to an Oral course, while those weak in all skills were assigned to the Integrated Language Skills course. Experience showed, however, that arbitrary decisions often had to be taken as to the course most suitable for individual students as there was rarely a clear-cut distinction between those weak in one or two skills as opposed to all four skills. On the basis of this finding, and the recognition that students, however good they may be in English, may still require English enhancement to maximize the benefits they accrue from their university studies, the Language Centre decided to streamline the courses it offered the Faculty of Arts students. At present it runs one course known as the English for Arts Students (EAS) course. This development is in line with current university thinking on English enhancement, which not only supports the extension of English programmes but is also moving towards making such courses credit-bearing. However, it weakens the case for testing on entry to university since the results are no longer needed for placement, but only for determining those students who are eligible for exemption from the course.

**Comparison of Tests in the Study**

In order to see whether the two tests under consideration, that is the UE and LAS, were indeed comparable, a detailed study of each subtest was undertaken. This, as can be seen from the brief descriptions of the subtests below, revealed that there appears to be considerable overlap across the subtests and the examinations as a whole, although there is no one-to-one correspondence between the subtests. There was therefore a sound basis upon which to proceed with the study.

**The New 'Use of English' Examination**

A fundamental change in testing philosophy underpinned the design of the new UE examination. Task authenticity became a primary consideration and is reflected in all four sections of the examination.

**Section A–Listening**

The Listening Paper sets out to test 'the ability of candidates to understand, organize and interpret spoken English as used by educated and fluent speakers of English as an international language' (HKEA, 1985: 2). The brief specifies that the spoken language used for the test is to be semi-formal rather than informal and that realistic contexts are to be provided. Furthermore, the candidates are to use the information obtained from listening to accomplish specific tasks and the test is to reflect real-life listening experiences. For example, candidates are allowed to listen once only to the information, but they are allowed to familiarize themselves with the task at hand before listening to the text so as to listen with a focus and purpose. An attempt is also made to cover a wide range of listening enabling skills, such as identifying main themes and supporting details, extracting global information and recognizing speakers' attitudes. The 1989 paper included the presentation of a monologue and dialogue conducted in a business context, while the 1990 paper used a monologue conducted in the form of a semi-academic lecture.

**Section B–Writing**

This test sets out to assess the candidates' ability to write extended discourse of an expository nature, e.g. in the form of persuasive arguments, short reports, development of

hypotheses, etc. rather than straight narratives or descriptions. Candidates are given a choice of four writing topics from which they select one on which to write an essay of about 500 words. Care is taken to give each topic a realistic context and the writer a purpose for which to write. Assessment is based on organization of ideas and coherence of arguments as well as accuracy and range of the language used.

**Section C--Reading and Language Systems**

The reading part of this test, which carries about a quarter of the weighting of this section, is aimed at testing 'the ability of candidates to achieve an "in depth" understanding of an expository text' (HKEA, 1987: 3). The emphasis is not on fact-retrieval but on global comprehension of the overall structure of the text, which in turn depends on the candidates' ability to recognize the purpose of the text as well as the attitude of the writer. Items set on lexis require candidates to deduce contextual meanings through adequate interpretations of the text.

The language systems part, which carries three-quarters of the weighting of this section, sets out to test

> the extent to which the systems of the English language have been internalized by the candidates:
> - the lexicon
> - the morphology
> - the syntactic relationship within and among phrases and clauses, and
> - the structural relationships among sentences within paragraphs and in discourse (HKEA, 1987: 3)

This paper, which is to a large extent objectively marked, represents an attempt to tap the candidates' general language proficiency in English using test-tasks set at discourse level. The tasks, which include summary cloze, proof-reading, paragraph-building, list-matching, etc., call for a good mastery of English lexis and syntax as well as an awareness of cohesion and coherence features in English.

**Section D--Practical Skills for Work and Study**

As the title of this section suggests, the paper requires candidates to display such practical English language skills as are needed in work and study situations. A bank of information in verbal and graphic form is given to candidates as a data-file. Candidates are required to use the given data to complete several specified communicative tasks. In this way candidates' abilities in processing, selecting and presenting relevant information in a fitting manner as specified by the tasks are tested.

> The emphasis will be upon practical rather than academic skills as such, though the techniques and skills tested are those which are required for professional work which requires the ability to process information, make judgements and formulate solutions to problems; they will thus be equally relevant to students entering tertiary education. (HKEA, 1987: 4)

**Language Analysis Sessions**

Although HKU Language Centre's LAS battery of tests has been described elsewhere (Fok, 1985) a brief description of all the subtests is given below for ease of comparison with the UE subtests. It is also worth noting that the subtests have undergone some changes over the last five years.

### 1. Listening

The aim of this subtest is to test the candidates' ability to follow the overall development of academic discourse and to identify relationships between main points and examples. The test takes the form of an unscripted 35-minute lecture which has been recorded on video. The candidates take notes while viewing/listening to the lecture which they see once only. They then use their notes to answer short questions.

The test is divided into four parts, respectively testing candidates' ability to: identify the main points of the lecture; correct certain inaccurate statements about the content of the lecture; recognize points made during the lecture by selecting those points from a list; identify the relationship between certain examples and the main points of the lecture.

### 2. Writing

This test is designed to test the candidates' ability to write a tutorial paper. It is divided into five sections and a common theme runs through them all. The candidates are given a certain amount of input in the form of notes of the type they are likely to collect from reference material. Each of the subsections tests a different enabling skill. Candidates are expected to proofread and correct grammatical errors within the introduction of a text; describe changes that have taken place over time; compare different theoretical views using information provided in the form of notecards; use appropriate connectives; reformulate sections of a text that have been marked in the margin as problematic in writing style.

### 3. Reading

This test is designed to test the candidates' ability to adopt appropriate reading strategies for dealing with a long piece of academic-type discourse. It tests the candidates' ability to understand the main ideas within different sections of the text as well as their understanding of specific information presented in the text.

### 4. Oral

This is a direct test of the candidates' ability to participate actively in a tutorial discussion. It takes the form of a simulated tutorial with a tutor leading a 15-minute discussion among four or five candidates. Before the actual discussion candidates are given time to read a short passage on an academic subject and they are given some questions to consider on the basis of the reading. During the discussion, the candidates are encouraged to interact with one another rather than merely respond to the tutor's comments and questions. They are assessed by the tutor as well as two other markers on the basis of the following criteria: fluency, accuracy and interactional skills.

## Procedure

The results for UE, LAS and end-of-first-year subject examinations were obtained for 346 first-year students admitted to the Faculty of Arts at HKU in September 1989. Details of these results are set out below.

### 1. UE Results

These consist of four weighted scores, one for each section of the UE paper, namely Listening, Writing, Reading & Language Systems, and Practical Skills for Work & Study. These weighted scores when added together make up the total UE score (henceforth UTOT), for which a UE grade is assigned to each candidate. The University of Hong Kong requires a minimum of grade D for admission to any of its programmes (In 1989, 46.3% of candidates sitting the UE exam achieved a grade D or above).

### 2. LAS Results

These are also weighted according to the four subtests, Reading and Listening carrying a weighting of 20%, and Writing and Oral a weighting of 30% each. They add up to an LAS total score (LTOT, henceforth).

### 3. B.A. First Year Examination Results

These are the results students received in their B.A. First Year Examinations in May 1990. In their first year students select 3 or 4 subjects from a choice of 17. Students' performance in these examinations determines their promotion to the second year of studies and their choice of major and minor subjects for their degree.

These results were statistically analysed to investigate:

a. Whether or not the new UE examination closely resembles the LAS tests in determining the subjects' overall language abilities on entry to university;

b. Whether or not the new UE examination taps the same language ability factors as the LAS tests;

c. Whether or not the new UE examination possesses a higher predictive ability than the LAS tests in terms of students' end-of-first-year results.

The following statistical analyses were carried out on the results:

a. UE total scores (UTOT) and LAS total scores (LTOT) were correlated to see whether the two tests yielded similar results. In addition, UE total scores (UTOT) were correlated against the LAS total minus the oral component (henceforth LTT), as oral skills are not tested in the UE examination (See table 1).

b. Factor analysis was carried out to see whether there were distinct UE and LAS language factors or whether specific factors overlapped across the two tests, that is the UE and LAS (See table 2).

c. UE and LAS results were correlated against the end-of first-year examination results to see which of the two tests performed as the better predictor of students' performance in their academic studies. In addition, correlations were run on the LAS total minus oral (LTT), since UE has no oral component (See table 3).



6

BEST COPY AVAILABLE

## Findings and Discussion

### 1. Correlation between the UE Totals and the LAS Totals

The correlation index between the UE total score and the LAS total score was .6117** (see table 1 below). Though statistically significant at .001 level, the correlation leaves almost two-thirds (64%) of the variance unaccounted for, and hence the significance is likely to be a result of the large sample in the study (n=346). However, an important point to note is that the population used in this study is more homogeneous in terms of level of proficiency than the total population sitting the UE examination, since all the candidates in the study had a minimum grade D(10) in UE, that is, they were taken from the top 40% of the total UE population.

Since the UE examination has no oral component, UTOT was also correlated against the total scores of the LAS Listening, Reading and Writing subtests (LTT). This was found to be marginally higher at .6444** indicating that the oral subtest of the LAS did not significantly affect students' overall performance. It further suggests that the differences between the two tests cannot be accounted for by the lack of an oral subtest in the new UE examination.

Table 1: Correlation Indices for UE and LAS Total Scores

| Correlations | UTOT | LTOT | LTT |
|---|---|---|---|
| UTOT | 1.0000 | .6117** | .6444** |
| LTOT | .6117** | 1.0000 | .8986** |
| LTT | .6444** | .8986** | 1.0000 |

1-tailed signif: ** - .001

### 2. Factor Analysis

To obtain further evidence of whether the UE and LAS were tapping the same subskills, and whether higher correlations exist between subtests of similar skills than between the complete battery of tests, factor analysis was carried out. Factor 1 came out as a very strong factor accounting for 35.5% of the total variance, whereas Factor 2 was much less significant, accounting for only 4.7% of the total variance. The individual factor loadings for the eight subtests are shown in Figure 2 below.

When we look at the factor loadings for Factor 1, the strongest factor by far, we find no evidence of distinct UE and LAS factors. This is contrary to the findings in the 1985 study (Fok, op cit), and suggests that the present tests are not distinctly different in nature as was suggested of the old UE and the LAS in the 1985 study. There is also no clear evidence to suggest that similarity of language skills yielded significant overlap in the variance of the 8 subtests. Therefore one might conclude that Factor 1 is an 'integrated language skills' factor, probably with slightly greater dependence on the reading and listening skills since the highest factor loadings are for LR, UR, LL and UL.

The highest factor loadings for Factor 2 are for LW (.62) and UW (.60) suggesting a writing factor over and above the integrated skills factor. However, since this factor only accounts for 4.7% of the total variance, the role played by writing per se in the two test batteries should not be overestimated.

Table 2: Factor analysis of subtests of UE and LAS

| | Factor 1 | Factor 2 |
|---|---|---|
| UL | .53 | .24 |
| UW | .10 | .60 |
| UR | .60 | .49 |
| UWASPS | .37 | .10 |
| LR | .73 | .19 |
| LW | .45 | .62 |
| LL | .57 | .22 |
| LO | .34 | .34 |

Key: UL - UE Listening
UW - UE Writing
UR - UE Reading & Language Systems
UWASPS - UE Work & Study Practical Skills
LR - LAS Reading
LW - LAS Writing
LL - LAS Listening
LO - LAS Oral

### 3. Predictive Abilities of the UE Exam and the LAS Tests

From the correlation indices given in Figure 3 below, it can be seen that the LTT scores have the strongest predictive power, followed by the LTOT scores. If we take the B.A. First Year Examination as the criterion behaviour then it would appear that the LAS test results have more in common with the target performance than do the results of the UE examination. However, this difference in predictive power is much less marked than that found between the LAS and the old UE examination in the 1985 study (Fok, op cit). In addition, one must recognize the fact that there are many factors which contribute to the B.A. Examination results and that it would be impossible to identify them all so as to isolate the proportion of variance attributable to English abilities alone.

If one looks closely at the ranking of individual correlation indices in each column, it can be seen that the LAS results conform more closely to our expectation as to which subjects require a higher level of proficiency in English. The LTT scores boast a highly significant .61** correlation with Comparative Literature, and a significant .47** correlation with English Language and Literature. The fact that significant correlations were obtained for such subjects as Political Science, Psychology, Geography and Geology, and History also offers no surprise as these are subjects which require students to write reasonably long English prose of a discursive nature. The fact that the LTT results produced a significant negative correlation of .27* against Chinese Language and Literature also seems consistent with the common-sense assumption that one does not need proficiency in English to do well in Chinese.

The LTOT results yielded a similar rank order of subject correlations to LTT, again with Comparative Literature and English Language and Literature correlating most highly and Chinese Language and Literature correlating negatively.

The UTOT results, by comparison, yielded less convincing figures with what may appear a puzzlingly high correlation of .59* (significant at the .01 level) for Statistics, the highest of all the subjects. This high correlation may, however, have been a result of the small number of subjects who took Statistics in the Faculty of Arts (n=20). The fact that English Language and Literature and Comparative Literature still come rather high on the list while Chinese Language and Literature and Computer Science are found at the bottom with negative correlation indices restores substantial credibility to the predictive power of the new UE examination.

Table 3: Correlations Indices between UE Exam Results, LAS Test Results and First Year Exam Results

| UTOT | | LTOT | | LTT | |
|---|---|---|---|---|---|
| ST | .59* | CL | .65** | CL | .61** |
| EN | .42** | EN | .49** | MU | .60 |
| FA | .39 | MU | .41 | EN | .47** |
| GG | .38** | JA | .36* | ST | .41 |
| CL | .34 | PS | .35** | PS | .38** |
| MU | .28 | ST | .28 | JA | .34 |
| JA | .24 | SY | .28* | FA | .30 |
| HI | .22 | FA | .27 | PY | .29** |
| PS | .20 | GG | .24* | GG | .28* |
| PH | .19 | CS | .23 | HI | .26* |
| SY | .18 | PY | .22* | SY | .15 |
| GM | .16 | HI | .21 | PH | .13 |
| PY | .18 | PH | .18 | EC | .07 |
| EC | .02 | EC | .09 | CS | .06 |
| CS | -.21 | GM | -.14 | GM | -.16 |
| CH | -.22 | CH | -.25* | CH | -.27* |

1-tailed signif: * - .01 ** - .001

Key:

| | | | | | |
|---|---|---|---|---|---|
| ST | - | Statistics | n | = | 20 |
| EN | - | English Language & Literature | n | = | 148 |
| FA | - | Fine Art | n | = | 15 |
| GG | - | Geography & Geology | n | = | 106 |
| CL | - | Comparative Literature | n | = | 41 |
| MU | - | Music | n | = | 13 |
| JA | - | Japanese | n | = | 45 |
| HI | - | History | n | = | 86 |
| PS | - | Political Science | n | = | 120 |
| PH | - | Philosophy | n | = | 89 |
| SY | - | Sociology | n | = | 89 |
| GM | - | German | n | = | 14 |
| PY | - | Psychology | n | = | 178 |
| EC | - | Economic | n | = | 145 |
| CS | - | Computer Science | n | = | 16 |
| CH | - | Chinese Language & Literature | n | = | 110 |

TotalY N=346 (students take 3 or 4 subjects in their first year at university)

## Conclusion

The above analyses seem to suggest that for the purpose of predicting prospective undergraduates' academic performance the LAS tests continue to function more effectively than the UE examination. This is not surprising if one bears in mind the fact that the LAS tests are targeting a much more homogeneous population, la crême de la crême both in terms of English proficiency and overall academic achievement (Although the LAS tests are a better predictor of undergraduates' first-year performance in their academic subjects, in some instances it would appear that the LAS tests would serve this purpose better if the oral subtest were removed from the test battery).

Despite there being considerable overlap in skills tested, and there being a possible common 'integrated skills' factor, the UE examination and LAS tests are not identical and could in no way be regarded as parallel. In order to maximise the information upon which decisions can be made about students' needs for English enhancement, it would seem necessary to continue with some form of internal pre-course testing. However, before any decisions can be made as to whether to retain an internal measure, one must consider the purpose of testing to see how far there is a real and justifiable need for the additional information provided by the LAS (or any equivalent) tests.

Our study suggests that the new UE examination appears to be more valid than the old UE both in terms of face and predictive validity. It may be a sufficient measure in cases where English enhancement is not equated with remediation, but is considered an integral part of English-medium tertiary education to be offered to all incoming students.

REFERENCES

Fok, Angela C. Y. Y. 1985. 'Language Proficiency and Related Factors'. In Lee, Y. P., Fok, Angela C. Y. Y., Lord, Robert & Low, Graham (eds.) *New Directions in Language Testing*, pp. 127-135. Oxford, Pergamon Press.

Hong Kong Examinations Authority 1987. *New Use of English Examination,* Hong Kong.